DOCUMENT RESUME

ED 392 848 TM 024 696

AUTHOR Shimberg, Benjamin
TITLE Assuring the Quality of Licensing and Certification Programs.
PUB DATE 21 Apr 95
NOTE 11p.; Presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21, 1995).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Certification; *Licensing Examinations (Professions); *Quality Control; *Standards; *Test Construction; Test Use
IDENTIFIERS High Stakes Tests; *Monitoring; Standards for Educational and Psychological Tests

ABSTRACT

When one considers the importance and social significance of licensing and certification examinations, it is amazing that the enterprise operates with virtually no societal oversight. The "Standards for Educational and Psychological Testing" and the "Code of Fair Testing Practices in Education" of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education are statements of ideals, but they lack any enforcement mechanism. Licensing and testing organizations are motivated to cut costs and show a profit, and their test development procedures are not always apparent to the test user. The creation of some sort of watchdog agency to monitor high-stakes tests might be advisable. The "Standards for Quality and Fairness in Testing" developed by the Educational Testing Service for its own use might be applicable to testing as a whole. These standards are enforced at ETS through a "Visiting Committee" that monitors the test auditing program. These standards may be a model for an external agency for the testing community that could conduct spot checks and grant approval of test development practices. Such an agency would be good for the testing community and for the public. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 392 848

ASSURING THE QUALITY OF LICENSING AND CERTIFICATION PROGRAMS

Benjamin Shimberg
Educational Testing Service

This audience does not need to be reminded of the importance of licensing and certification tests. All of us understand how much is at stake each time one of these tests is administered

For the individual test-taker, it may represent the culmination of years of effort to prepare to enter or gain a needed credential for engaging in a regulated occupation.

For the credentialing agency, it represents a crucial step in the qualification process.

"Does the applicant meet the standards set for licensing or certification." Unless the evaluation instruments are valid and reliable, the entire process may be flawed.

And finally, there are the consumers--the users of regulated services. They have a great deal at stake.If the licensed provider is not qualified, their health and safety...or perhaps their financial security...could be in jeopardy.

When one considers the importance...indeed, the social significance of licensing and certification exams...it is truly amazing how the entire enterprise operates WITH VIRTUALLY NO SOCIETAL OVERSIGHT.

There are, of course, the standards for Educational and Psychological Testing promulgated by APA, AERA and NCME. These standards have been evolving and changing since the early 1950s to meet the ever-changing needs of the psychometric community.

There is also the Code of Fair Testing Practices in Education developed by a joint committee of these same organizations.

These are great, as far as they go. However, one needs to be aware of the political infighting that occurred, especially in connection with the Standards.

Prepared for presentation at AERA/NCME symposium, Licensure and Certification: Protecting Examinees, the Public, and Decision Makers.San Francisco. April 21, 1995

TM024694

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
BENJAMIN SHIMBERG
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

There's a lot of deliberate ambiguity.

More often than not, the Standards are statements about ideals..... rather than statements of the realities of the World as we know it in 1995.

Moreover, the Standards lack any sort of enforcement mechanism.

I should also add that the 1985 Standards have placed heavy reliance on professional judgement in interpreting individual standards.

I have no quarrel with professional judgement,

But I'm inclined to agree with my colleague, Sam Messick, who wrote in 1988, "...In the absence of enforcement mechanisms, where is the protection against unsound professional judgement?

Messick has put his finger on what is probably the single greatest shortcoming of the Joint Standards....the absence of any enforcement...or even a monitoring mechanism.

Over the years, psychologists and educational researchers involved in testing enterprises, have adopted a "see no evil, speak no evil" attitude toward testing in general...and high stakes testing in particular.

Many of them, I'm sure, would like to believe that we still operate in a bygone era

when psychological testing was primarily a cottage industry, conducted by academics who thought of testing along the lines of the experimental psychology model.

You know the good old days when the "null hypothesis" was the coin of the realm among psychologists and statisticians and where cutting corners was unthinkable

But all that changed after the end of World War 2.

Testing has proved its worth as a tool for the screening and placement of recruits.

The tremendous success of the Air Force Aviation Psychology program elevated testing to the status of a secret weapon. Small wonder that industry seized on this technology to take much of the guesswork out of personnel selection.

Testing also came to play a critical role in admissions to colleges and professional schools, in the assessment of educational achievement and in educational and vocational guidance.

Needless to say, licensing and certification agencies quickly abandoned their old-fashioned essay exams in favor of the more objective, easier to score multiple choice tests.

As testing moved from cottage industry to big business, competitive pressures developed. Small, start-up companies were eager to get contract.

Larger testing organizations had substantial investments in staff and equipment. They needed lots of business to keep these assets productively employed.

In such a competitive climate, it would not be unusual for companies to bid low...in order to get the business....and then look for ways to cut costs in order to show a profit.

To a client..especially one who is not psychometrically sophisticated...one test looks pretty much like another. It's impossible to tell...just by looking at the questions...what procedures actually went into the development process.

Take job analysis, for example. There is no general agreement regarding how large a sample should be....or what needs to be done to assure that it is representative of the target population.

One can skimp on the job analysis effort and the client will seldom be the wiser. If challenged, the contractor can always say that in his or her professional judgement the sample was adequate.

One could make similar observations about the writing and review and keying of test questions.

About the depth of the statistical analysis to identify faulty of miskeyed questions.

How much time and effort to invest in scaling and equating are also matters of professional judgement.

Some testing organizations tell clients that such procedures are unnecessary frills. Why spend money on such refinements? Candidates will never know the difference?

Then there is the matter of screening questions for possible gender or ethnic bias. This,too, is a quality procedure that is not visible to the test-taker or to the client.

> Setting passing points can be another contentious area. There are different approaches. Which one to use is a matter of professional judgement. But are they all equally defensible?

As I indicated earlier, the Joint Standards are vague on many points and therefor provide little in the way of guidance for the test developer...

> ....or for the client who wants to be sure that he or she is getting a quality examination,

The secrecy which surrounds so much of the testing enterprise poses special problems for test takers.

> How can candidates be sure that the "high-stakes" tests they are taking meet professional standards?
>
> > Even if candidates could gain access to tests and other relevant documents, they would not be able to interpret the information or make informed judgements.

This suggests than an alternative approach may be needed...perhaps the creation of some sort of surrogate agency that would act as a watchdog to monitor "high stakes" tests on behalf of test takers and the public.

> Back in 1987, when testing organizations were embroiled in another "Truth in Testing" controversy, I wrote and article suggesting that the testing industry ought to create some sort of OVERSIGHT MECHANISM....
>
> > .... to assure that all tests offered for public use were of high quality and met professional standards.

I predicted that if the industry, itself, did not assume oversight responsibilities, there was good likelihood that either the states or the Federal government might some day do so.

> I won't go into detail about the many ways in which the idea got kicked around over the next few years.
>
> > Suffice to say, that along the way I discovered that my concerns about the lack of oversight were shared by Dr. George Madaus, Director of the Center Study of Testing, Evaluation and Public Policy at Boston College.

While Dr. Madaus was interested primarily in the impact of tests in the educational world, he recognized that the same problems applied to employment tests and licensing and certification exams.

So we joined forces. I helped Dr. Madaus obtain a grant from the Ford Foundation and the Carnegie Corporation to conduct a study--at Boston College--

of the feasibility of establishing some sort of independent agency to monitor the quality of tests in both the educational world and the world of business and professional regulation.

All this happened about the time when I was getting ready to retire from ETS. After that, I lost touch with what Dr. Madaus was doing.

Several years later, when I received a copy of his report, A Proposal for a Monitoring Body for Tests Used in Public Policy.

I found that he had decided to restrict his study to the fields of elementary and secondary school testing.

Moreover, he had concluded that what was needed was a Board that would intercede in testing controversies only when invited to do so

For example, he says...

"...The Board might receive requests for monitoring from a variety of sources, such as state departments of education, advocacy groups, test developers, or sponsors, or individuals.

All such requests would serve as triggers for consideration of monitoring or auditing by the board" (99)

While the Board would use as its starting point the technical and ethical standards developed by professional groups, he believes that

"...the Board should not attempt generically to operationalize and apply these standards.

Instead, the Board should negotiate each standard in light of the particular context of use and audience and seek to promote consensual agreement around how the standard is to be applied".(p 100)

There isn't time to go into detail about the full scope of Dr. Madaus' proposal. Suffice to say that..in my view.. his proposal fails to address the concerns I voiced in 1987 and which are still germane today.

So I've decided that it may be time to put back on the table an idea that I believe may have merit for assuring the quality of tests.

Perhaps the best way to explain what I have in mind is to describe, briefly, a model that has been in use at ETS for about 15 years.

Under the leadership of Bill Turnbull, who was President of ETS at the time, the staff developed and the ETS Board of Trustees adopted, a set of Principles and Guidelines for Quality and Fairness in Testing These were,subsequently, renamed The ETS Standards for Quality and Fairness in Testing.

These were,,and still are... congruent with the APA, AERA, NCME Technical Standards. However, they are more detailed and more explicit.

If you'd like a copy of the ETS Standards for Quality and Fairness in Testing,just write to ETS for a copy.

All ETS testing programs are required to adhere to the Standards.

To insure compliance, an Office of Corporate Quality Assurance was established. This office conducts periodic audits of all ETS programs.

The actual reviews or audits are done by members of the ETS professional staff...none of whom have had any involvement with the program being audited.

There are also five EXTERNAL AUDITORS--Testing experts from outside of ETS--who work with the audit teams and then write confidential reports of their observation. These reports go directly to the Visiting Committee, which I'll describe in a moment.....

Since members of the audit teams have no direct involvement with the program under review, they can be highly critical..sometimes even brutally so...when circumstances warrant.

These staff members--and the external auditors-- receive training in the audit procedure and are then provided with the back-up materials they need in order to judge how well the specific program under review meets the ETS Standards.

These materials are assembled, beforehand, by the Program Manager in accordance with a procedural guide, which details what documentation is needed.

The audit team reviews the documentation, asks many questions and decides on the extent to which the program is or is not in compliance with the ETS Standards.

The Manager of the program being audited is given an opportunity to review and comment on a preliminary draft of the auditor's report.

However, after an exchange of views, the audit team sends its final report to the Cognizant Vice President, who may then have discussions with the Program Manager to set a time table for overcoming any deficiencies that were identified.

A skeptic might say, "All that sounds pretty good, but how can one be sure that it really happens that way?

What if it is just a paper program concocted for public relations purposes?"

To make sure the program operates as intended and is not just window dressing, the ETS Board of Trustees has established a Visiting Committee of sixteeen distinguished people whose job it is to monitor the audit program to make sure it really works.

The Committee is made up of outstanding people from the fields of measurement and education as well as from the public at large, including critics of testing...such as the Executive Director of the NEA.

The Visiting Committee spends several days at ETS reviewing the audit process in-depth.

The Committee reviews all the programs that were audited over the past year .

They talk with the program staff, with members of the audit team, and with others who may be involved in the program...such as the researchers who may have been responsible for the job analysis.

They also review the confidential reports submitted by the EXTERNAL AUDITORS, who worked with the various audit teams.

The goal of the VISITING COMMITTEE is to assure itself regarding the integrty of the audit process.

They also check to see if recommendations made in previous years have been implemented.

The Visiting Committee then prepares a report which goes directly to the ETS Board of Trustees, detailing any observed shortcomings in the audit process and suggesting any needed improvements.

If you'd like to see one or more reports of the VISITING COMMITTEE, they are available on request FROM ETS

While this model serves the needs of ETS very well, there is no assurance that it would fit the needs of other testing organizations.

However, that was not my purpose in describing the ETS model. I have described it in some detail because I believe it may provide a starting point for thinking about ways to impose some social controls within the testing industry where none exist at the present time.

What I have in mind is a VOLUNTARY effort... on the part of the testing community...

First to see if basic principles and guidelines...fair to both large and small testing organizations...can be developed, including agreement on what constitutes reasonable documentation of compliance.

I would then encourage testing organizations to institute internal audit procedures. Such audits would serve to educate staff members regarding the standards and the importance of good documentation.

After the industry had an opportunity to institute audit procedures,

I'd like to see an external agency created to conduct spot checks--similar to those done by the ETS Visiting Committee-- to determine whether the audit procedures were working as intended.

Organizations whose audits passed muster could be certified as adherents to sound practice standards.

Certification would be the equivalent of the Good Housekeeping seal of approval, or the UL designation given to electrical appliances by the Underwriter's Laboratory.

At this point, many of you may be wondering "Why, in the absence of any legal requirement, would any testing organization agree to have its procedures audited by an outside group?"

They might do so if they knew that prospective clients-- such as licensing and certification agencies, groups that conduct national testing programs for various professional groups, state departments of education, and other large-scale users of high-stakes tests--

would not do business with any organizations that had not been certified as conducting internal audits to assure compliance with the agreed-upon standards.

Testing organizations that are already following good practices would find it relatively easy to install internal audit procedure.

Those with less rigorous standards would be motivated to shape up if they wanted to compete for large-scale testing contracts.

What appeals to me about this process is that once it is in place, NOT following sound procedures would have CONSEQUENCES...for the individuals involved and for the organization.

Both would have incentives to think about what it would take to install appropriate remedies.

I recognize that the process I am proposing isn't foolproof. I am not suggesting that every test be checked out before being put into operational use.

That might be politically correct, but not very practical.

Sure it is possible that an organization that generally follows good practices may, on occasion, turn out a defective test. This could happen.

But it is less likely to happen here than in situations where there is no oversight at all.

There are,of course many unanswered question.

Will it be possible to get agreement on what constitutes acceptable or desirable professional practice?

If that hurdle can be surmounted...if we can operationalize what we mean by good practice...and if we can get testing organizations to commit themselves, voluntarily, to documenting their compliance..

...then we can move on to the next step, which would be the creation of a mechanism to insure that the agreed upon standards are, indeed, being followed.

I'm not suggesting that the task will be easy.

On the contrary, it will be hard work.

Nevertheless, I'd like to see an effort made to tackle the problem, rather than just sitting around doing nothing because we are unable to achieve perfection.

I believe that such an effort will be

good for the testing community;

good for those who use tests to make critical decisions;

good for test-takers, whose lives are so often influenced by the tests they are required to take; and

good for the public, which often relies on licensing and certification to protect it from unqualified providers.

Thank you.